



Penentuan Kelayakan Karyawan Baru Menggunakan Data Mining Dengan Algoritma *Decision Tree* (C4.5)

Hamzah Kadar^{1*}, Agus Budiyantra²

¹⁻² STMIK Widuri, Indonesia

hamzah23412053@kampuswiduri.ac.id^{1*}, agusbudiyantra@kampuswiduri.ac.id²

Alamat: Jl. Palmerah Barat No.353, RT.3/RW.5, Grogol Utara, Kby. Lama, Kota Jakarta Selatan, DKI Jakarta 11480.

Korespondensi penulis: hamzah23412053@kampuswiduri.ac.id*

Abstract. Eligibility for new employees includes individuals who have skills appropriate to the position they are applying for, have a high willingness to learn, communicate well, and have integrity and good work ethics. They must also be able to adapt to the work environment and team quickly, but determining the suitability of new employees is quite difficult given the competencies of each division, therefore the use of data mining is very suitable for determining the suitability of new employees according to the needs of the company which uses them. *decision tree algorithm* (C4.5), the results obtained from the *decision tree algorithm* process show the truth tree for classifying new employees and a high level of accuracy with a percentage of 98.44% based on test 2.

Keywords Employee, Data Mining, C4.5

Abstrak. Kelayakan karyawan baru mencakup individu yang memiliki keterampilan yang sesuai dengan posisi yang mereka lamar, memiliki kemauan belajar yang tinggi, berkomunikasi dengan baik, serta memiliki integritas dan etika kerja yang baik. Mereka juga harus mampu beradaptasi dengan lingkungan kerja dan tim dengan cepat, namun penentuan kelayakan karyawan baru yang dirasa cukup sulit dengan kompetensi terhadap divisi masing – masing, oleh sebab itu penggunaan data mining sangat cocok untuk penentuan kelayakan dari karyawan baru sesuai kebutuhan perusahaan yang dimana menggunakan algoritma *decision tree* (C4.5), hasil yang diperoleh pada proses algoritma *decision tree* menunjukkan pohon kebenaran klasifikasi karyawan baru dan tingkat akurasi yang tinggi dengan persentase 98,44% berdasarkan pengujian tes 2.

Kata kunci: Karyawan, Data Mining, C4.5

1. LATAR BELAKANG

Karyawan baru merupakan individu yang baru saja bergabung dengan sebuah organisasi atau perusahaan yang membawa energi segar dan antusiasme yang tinggi dalam menyesuaikan diri dengan lingkungan kerja baru diperusahaan, umumnya karyawan baru memiliki semangat untuk belajar dan berkembang serta menunjukkan komitmen yang kuat terhadap pekerjaan mereka. Selain itu karyawan baru perlu mendapatkan arahan dan dukungan dari rekan kerja dan atasan mereka untuk membantu mereka mengintegrasikan diri dengan tim dan memahami tugas-tugas mereka dengan baik. Proses adaptasi ini penting bagi mereka agar dapat berkontribusi secara maksimal dan menjadi anggota tim yang berharga (Hidayah & Rozi, 2021; Husein & Brutu, 2022; Rudianto et al., 2022; Saputra et al., 2023)

Peranan karyawan baru diperusahaan untuk membantu mencapai tujuan perusahaan yang diharapkan untuk belajar tentang budaya dan nilai-nilai perusahaan, serta memahami

tugas dan tanggung jawab individu dengan baik. kemudian mengikuti arahan atasan dan bekerja sama dengan tim sehingga dapat berkontribusi pada kesuksesan perusahaan melalui ide-ide baik, kerja keras, dan dedikasi. Disamping itu sebagai karyawan baru memiliki kesempatan untuk membawa perspektif baru dan inovasi ke meja, membantu perusahaan beradaptasi dengan perubahan lingkungan bisnis yang terus berkembang. Dengan memperlihatkan motivasi, kemauan untuk belajar, dan sikap yang positif, Anda dapat membangun reputasi yang kuat dan menjadi aset berharga bagi perusahaan (Adria, 2022; Khoeri & Iskandar Mulyana, 2021; Leidiyana & Permana, 2020; Romli & Zy, 2020)

Berdasarkan aspek karyawan baru terhadap persoalan kelayakan diperusahaan pada divisi masing – masing tentunya mengalami kesulitan dalam penentuan kompetensi dari individu, oleh karena itu pemecahan masalah yang dialami menggunakan algoritma *decision tree* (C4.5), algoritma *decision tree* ialah konsep data mining yang membangun pohon keputusan untuk klasifikasi dengan menggunakan aturan "splitting" berbasis informasi untuk membagi data menjadi subset yang homogen. Lalu setiap simpul dalam pohon mewakili keputusan berdasarkan fitur-fitur data. C4.5 dapat menangani data kategorikal dan numerikal serta mengatasi masalah "overfitting" dengan pruning (Abdul Fath Ramadhan, 2024; Rahmawan, 2020; Tulus et al., 2022)

Adapun yang menjadi tujuan penelitian untuk mengetahui mengukur kelayakan karyawan baru sesuai kebutuhan perusahaan dari proses algoritma *decision tree* (C4.5), namun hasil yang akan diperoleh berupa pohon keputusan dalam penentuan tersebut.

2. KAJIAN TEORITIS

Data mining menjadi proses penemuan pola, hubungan, atau informasi bermanfaat lainnya dari kumpulan data yang besar dan kompleks. Tujuan utamanya adalah mengidentifikasi informasi tersembunyi yang relevan dan berguna untuk pengambilan keputusan. Proses ini melibatkan penggunaan metode statistik, matematika, dan kecerdasan buatan, seperti algoritma pembelajaran mesin, untuk menganalisis data. Data mining umumnya diterapkan pada berbagai bidang, seperti pemasaran, keuangan, kesehatan, dan manufaktur, untuk membantu dalam memprediksi tren, perilaku konsumen, deteksi anomali, serta pengembangan strategi bisnis. Prosesnya terdiri dari beberapa tahapan, termasuk pembersihan data, integrasi data, seleksi fitur, transformasi data, serta penambangan pola atau model. Data mining sering dikaitkan dengan big data, karena pertumbuhan eksponensial volume data yang memerlukan teknik analisis canggih. Dengan kemampuan mengolah data secara efisien, data

mining menjadi alat penting dalam eksplorasi pengetahuan dari data yang tersebar dan tak terstruktur (Abidin et al., 2023; Asri et al., 2022; Putri et al., 2021).

Algoritma *decision tree* (C4.5) mencakup konsep data mining yang digunakan untuk klasifikasi dan prediksi. Algoritma ini membangun pohon keputusan berdasarkan atribut-atribut dari dataset untuk memprediksi label kelas dari sebuah data. Pada C4.5, pohon keputusan dibentuk dengan memilih atribut yang memberikan keuntungan terbesar dalam membagi dataset berdasarkan konsep *Information Gain* yang dihitung dari entropi. Atribut dengan nilai *Information Gain* tertinggi akan dipilih sebagai *node*, kemudian *dataset* akan dipecah menjadi subset berdasarkan atribut tersebut. Proses ini diulangi secara rekursif untuk setiap subset sampai semua data dalam subset memiliki kelas yang sama atau tidak ada atribut yang tersisa (Purnama et al., 2020; Rudianto et al., 2022).

3. METODE PENELITIAN

Proses *Knowledge Discovery in Database* (KDD)

Pada tahapan penelitian ini terdapat data set pelamar dari divisi HRD untuk sampel total 960 orang yang memiliki kategori divisi perusahaan. berikut data divisi bagi calon pelamar yang dapat dijabarkan dibawah ini

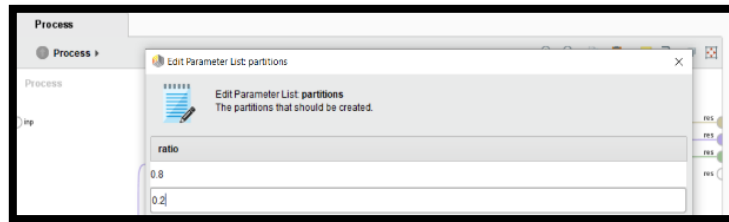
Tabel 1 Kategori Divisi Pelamar Kerja

Posisi Dilamar	Jenis Kelamin	Pendidikan	Pengalaman Kerja	Tes 1	Tes 2	Hasil
Customer Service	Laki-laki	S1	Tidak Berpengalaman	Kurang	Baik	Diterima
Admin	Perempuan	S1	Tidak Berpengalaman	Baik	Baik	Diterima
Customer Service	Laki-laki	SMA	Tidak Berpengalaman	Cukup	Kurang	Tidak Diterima
Customer Service	Perempuan	SMA	Tidak Berpengalaman	Kurang	Kurang	Tidak Diterima
Customer Service	Perempuan	SMA	Tidak Berpengalaman	Cukup	Cukup	Tidak Diterima
Customer Service	Perempuan	S1	Tidak Berpengalaman	Baik	Cukup	Diterima
Admin	Perempuan	SMA	Tidak Berpengalaman	Baik	Baik	Diterima
Customer Service	Perempuan	SMA	Tidak Berpengalaman	Baik	Baik	Diterima
Admin	Perempuan	SMA	Tidak Berpengalaman	Baik	Cukup	Diterima

4. HASIL DAN PEMBAHASAN

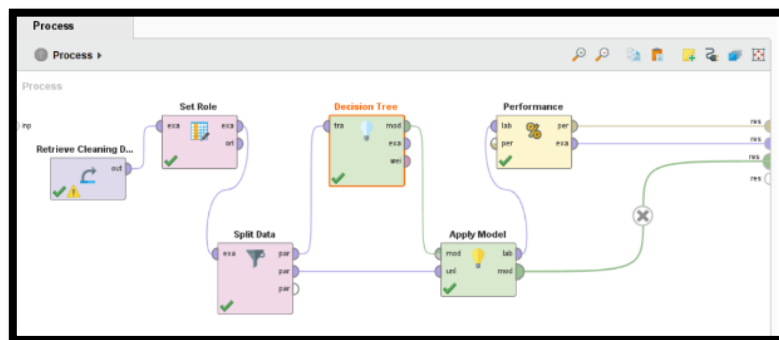
Proses Operator dan *Split Data Rapid Miner*

Tahapan proses operator dan *split data* melalui *rapid miner*, dilakukan awal tahapan operator terlebih dahulu, berikut operator dalam *rapid miner* antara lain:



Gambar 1 Operator *Rapid Miner*

Dari gambar diatas total data yang digunakan adalah 960 data pelamar, perhitungan dengan angka 0.8 atau sebanyak 768 data digunakan sebagai data training dan perhitungan angka 0.2 atau sebanyak 192 data digunakan sebagai data testing.

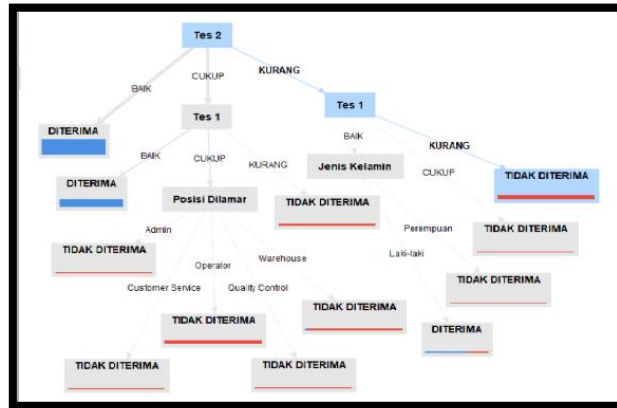


Gambar 2 Proses *Rapid Miner*

Dari gambar diatas operator dalam analisis data dan pembelajaran mesin memiliki peran penting dalam proses pengolahan dan pemodelan data. Retrieve Data Set digunakan untuk memuat data ke dalam proses, sementara Set Role membedakan atribut sebagai koordinat atau label. Split Data membagi dataset menjadi bagian yang proporsional untuk pelatihan dan pengujian. *Decision Tree* adalah model yang dapat memproses atribut nominal dan numerik untuk klasifikasi atau regresi. Apply Model menguji data dengan algoritma pembelajaran mesin yang sudah diterapkan sebelumnya. Performance digunakan untuk mengukur kinerja model. Melalui serangkaian operator ini, data diproses, model dilatih, dan kinerja model dievaluasi untuk mendapatkan pemahaman yang lebih baik tentang data dan membuat prediksi yang akurat. Setiap langkah memiliki peran khusus dalam memastikan keberhasilan analisis data dan pembelajaran mesin (Abidin et al., 2023; Ananda et al., 2022; Gina Sonia & Indriyani, 2023; Siallagan et al., 2022)

Klasifikasi Algoritma *Decision Tree* (C4.5)

Setelah tahapan operator dan *split* data, memperoleh klasifikasi algoritma *decision tree* atau pohon keputusan, berikut klasifikasi algoritma *decision tree* antara lain:



Gambar 3 Pohon Keputusan Algoritma C4.5

Dari gambar diatas dapat dijelaskan beberapa poin mencakup klasifikasi kelayakan divisi diperusahaan bagi karyawan baru antara lain:

1. Jika (Tes 2) mendapatkan hasil Baik, maka hasil keputusannya adalah Diterima. Hasil yang didapati apabila nilai dari Hasil Tes 2 mendapatkan kategori Baik, maka perusahaan dapat menyatakan bahwasanya pelamar tersebut dapat Diterima pada perusahaan ini.
2. Jika (Tes 2) mendapatkan hasil Cukup dan (Tes 1) mendapatkan hasil Baik, maka hasil keputusannya adalah Diterima. Berdasarkan penilaian dari Hasil Tes 2 yang hanya mendapatkan nilai Cukup, maka perusahaan dapat melihat Kembali dari hasil Tes 1 nya, apabila hasil dari Tes 1 nya mendapatkan nilai yang Baik, maka pelamar masih dapat dikategorikan sebagai kandidat yang dapat Diterima pada perusahaan ini.
3. Jika (Tes 2) mendapatkan hasil Cukup dan (Tes 1) mendapatkan hasil Cukup, maka hasil keputusannya adalah Tidak Diterima. Melihat dari hasil penilaian Tes 2 yang mendapatkan nilai Cukup, maka perusahaan dapat melihat Kembali pada hasil dari Tes 1, jika hasil Tes 1 hanya mendapatkan nilai Cukup sama dengan nilai Tes 2, maka pelamar disini sudah dapat dikategorikan sebagai pelamar yang Tidak Diterima pada perusahaan tanpa dilihat kebalikan dari berbagai atribut lainnya yang digunakan oleh perusahaan sebagai penunjang dalam proses penerimaan karyawan baru ini.
4. Jika (Tes 2) mendapatkan hasil Cukup dan (Tes 1) mendapatkan hasil Kurang, maka hasil keputusannya adalah Tidak Diterima. Seperti Analisa sebelumnya, apabila hasil dari kegiatan Tes 2 hanya mendapatkan nilai yang cukup, maka perusahaan dapat melihat kembali dari hasil kegiatan pada Tes 1, namun apabila hasil Tes 1 masih mendapatkan

nilai yang Kurang dari harapan perusahaan, maka kandidat tersebut sudah dikategorikan sebagai kandidat yang Tidak Diterima oleh perusahaan.

5. Jika (Tes 2) mendapatkan hasil Kurang dan (Tes 1) mendapatkan hasil Baik, Dan Jenis Kelamin Laki-laki maka hasil keputusannya adalah Diterima namun jika Jenis Kelamin Perempuan maka Tidak Diterima. Sesuai dengan kebutuhan perusahaan, kegiatan dari proses tes ini, apabila hasil dari Tes 2 mendapatkan nilai yang kurang dari kriteria perusahaan, maka masih dapat dilihat berdasarkan hasil Tes 1, apabila hasil dari Tes 1 tersebut mendapatkan nilai Baik dan mengutamakan pada Jenis Kelamin Laki-laki dibandingkan Perempuan, maka kandidat yang berjenis Laki-laki tersebut masih dapat dipertimbangkan kedalam kategori kandidat yang dapat Diterima oleh perusahaan.
6. Jika (Tes 2) mendapatkan hasil Kurang dan (Tes 1) mendapatkan hasil Cukup, maka hasil keputusannya adalah Tidak Diterima. Apabila kandidat pelamar mendapatkan hasil Tes 2 Kurang dan hasil Tes 1 masih masuk kedalam penilaian Cukup, maka kandidat tersebut tetap masuk kedalam kategori Tidak Diterima pada perusahaan.
7. Jika (Tes 2) mendapatkan hasil Kurang dan (Tes 1) mendapatkan hasil Kurang, maka hasil keputusannya adalah Tidak Diterima. Jika hasil Tes 2 Kurang serta hasil dari Tes 1 masih Kurang, maka sudah pasti kandidat tersebut masuk kedalam kategori pelamar yang Tidak Diterima oleh perusahaan.

```
Tree
Tes 2 = BAIK: DITERIMA (DITERIMA=943, TIDAK DITERIMA=0)
Tes 2 = CUKUP
| Tes 1 = BAIK: DITERIMA (DITERIMA=165, TIDAK DITERIMA=0)
| Tes 1 = CUKUP
| | Posisi Dilamar = Admin: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=11)
| | Posisi Dilamar = Customer Service: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=4)
| | Posisi Dilamar = Operator: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=65)
| | Posisi Dilamar = Quality Control: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=13)
| | Posisi Dilamar = Warehouse: TIDAK DITERIMA (DITERIMA=1, TIDAK DITERIMA=31)
| Tes 1 = KURANG: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=32)
Tes 2 = KURANG
| Tes 1 = BAIK
| | Jenis Kelamin = Laki-laki: DITERIMA (DITERIMA=12, TIDAK DITERIMA=6)
| | Jenis Kelamin = Perempuan: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=1)
| Tes 1 = CUKUP: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=3)
| Tes 1 = KURANG: TIDAK DITERIMA (DITERIMA=0, TIDAK DITERIMA=81)
```

Gambar 4 Deskripsi Pohon Keputusan Algoritma C4.5

Dari gambar diatas menunjukkan deskripsi secara rinci terhadap pohon keputusan yang sudah dicantumkan.

Perhitungan Manual Algoritma *Decision Tree* (C4.5)

Hasil pohon keputusan dari algoritma *decision tree* menggunakan *rapid miner* sebelumnya, dapat diperhitungkan manual algoritma *decision tree* antara lain

Tabel 2 Nilai Entropy dan Gain

		Data	Yes	No	Entropy	Gain
Total		960	651	309	0.906408808	
Posisi						0.012318009
	Admin	50	25	25	1	
	Customer Service	75	43	32	0.984426898	
	Operator	398	270	128	0.906121577	
	Quality Control	152	101	51	0.920473223	
	Warehouse	285	212	73	0.820866077	
JK						0.000241969
	Laki-laki	750	512	238	0.901457948	
	Perempuan	210	139	71	0.922984305	
Pendidikan						0.000996667
	SMA	886	605	281	0.901261588	
	D3	14	8	6	0.985228136	
	S1	60	38	22	0.948078244	
Pengalaman Kerja						0.000799848
	Berpengalaman	695	478	110	0.895731108	
	Tidak Berpengalaman	265	173	92	0.931515027	
Tes 1						0.417152421
	Baik	506	496	217	0.140106913	
	Cukup	251	94	157	0.954066225	
	Kurang	203	61	142	0.88189223	
Tes 2						0.417152421
	Baik	418	418	0		
	Cukup	410	220	190	0.996134484	
	Kurang	132	13	119	0.464174194	

Klasifikasi data mining dengan algoritma C4.5 (Decision Tree) akan menggunakan sebuah data training sebanyak 960 data pelamar untuk dapat menentukan sebuah akar dari atribut yang digunakan, maka akan dilakukan sebuah perhitungan nilai Entropy, kemudian melakukan perhitungan nilai Gain untuk setiap atributnya dengan rumus dibawah ini.

$$Entropy(S) = -\sum_{i=1}^n p_i \cdot \log_2 p_i \dots\dots\dots(1)$$

$$Entropy \text{ total} = ((-651/960 * \log_2(651/960)) + (-309/960 * \log_2(309/960))) = 0.906408808$$

Sehingga dari perhitungan diatas, maka didapati hasil dari *Entropy* total = 0.906408808 Setelah itu masing – masing atribut ini ditentukan untuk dijadikan sebagai parameter data yang digunakan pada perusahaan untuk dapat menentukan kelayakan karyawan baru.

1. Atribut Posisi Yang Dilamar

Dalam atribut ini, terdapat 50 data untuk posisi Admin, 75 data untuk posisi Customer Service, 398 data untuk posisi Operator, 152 data untuk posisi Quality Control dan 285 data untuk posisi Warehouse. Sehingga nilai *Entropy* dari Posisi adalah sebagai berikut :

Entropy (Admin) :

$$=((-25/50)*\text{Log}_2(25/50)+(-25/50)*\text{Log}_2(25/50)) = 1$$

Entropy (Customer Service) :

$$= ((-43/75)*\text{Log}_2(43/75)+(32/75)*\text{Log}_2(32/75))$$

$$= 0.984426898$$

Entropy (Operator) :

$$= ((-270/398)*\text{Log}_2(270/398)+(-128/398)*\text{Log}_2(128/398))$$

$$= 0.906121577$$

Entropy (Quality Control) :

$$= ((-101/152)*\text{Log}_2(101/152+(-51/152)*\text{Log}_2(51/152))$$

$$= 0.920473223$$

Entropy (Warehouse) :

$$= ((-212/285)*\text{Log}_2(212/285)+(-73/285)*\text{Log}_2(73/285))$$

$$= 0.820866077$$

Kemudian dalam perhitungan nilai *Gain* dapat digunakan rumus:

$$Gin(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|Si|}{|S|} * Entropy(Si).....(2)$$

$$\begin{aligned} \text{Gain} & \quad \text{Posisi} & = & \quad (0.906408808)-((50/960)*1)- \\ & ((75/960)*0.9844269898)((398/960)*0.906121577)-((152/960)*0.920473223)- \\ & ((285/960)*0.820866077 \\ & = 0.012318009 \end{aligned}$$

2. Atribut Jenis Kelamin

Pada atribut ini memiliki 750 data Laki-laki dan 210 data Perempuan. Sehingga untuk nilai *Entropy* nya sebagai berikut :

Entropy (Laki-laki)

$$=((-512/750)*\text{Log}_2(512/750)+(-238/750)*\text{Log}_2(238/750))$$

$$= 0.901457948$$

Entropy (Perempuan)

$$= ((-139/210)*\text{Log}_2(139/210)+(-71/210)*\text{Log}_2(71/210))$$

$$= 0.922984305$$

Kemudian untuk menghitung nilai *Gain* Jenis Kelamin sebagai berikut :

$$\begin{aligned} \text{Gain (JK)} & = (0.906408808)-((750/960)*0.901457948)-((210/960)*0.922984305)) = \\ & 0.000241969 \end{aligned}$$

3. Atribut Pendidikan

Didalam atribut Pendidikan, memiliki 886 data untuk lulusan SMA, 14 data untuk lulusan D3 dan 60 data untuk lulusan S1. Sehingga untuk dapat mengetahui nilai *Entropy* nya adalah sebagai berikut :

Entropy (SMA) :

$$= ((-605/886)*\text{Log}_2(605/886)+(-281/886)*\text{Log}_2(281/886)) = 0.901261588$$

Entropy (D3) :

$$= ((-8/14)*\text{Log}_2(8/14)+(-6/14)*\text{Log}_2(6/14)) = 0.985228136$$

Entropy (S1) :

$$= ((-38/60)*\text{Log}_2(38/60)+(-22/60)*\text{Log}_2(22/60)) = 0.948078244$$

Kemudian untuk mendapatkan hasil nilai *Gain* dari atribut Pendidikan, maka dapat menggunakan perhitungan sebagai berikut :

$$\text{Gain Pendidikan} = (0.906408808)-((886/960)*0.901261588)-((14/960)*0.985228136)-((60/960)*0.948078244) = 0.000799848$$

4. Atribut Pengalaman Kerja

Berdasarkan atribut Pengalaman Kerja ini, memiliki data sebanyak 695 untuk data berpengalaman dan 265 data untuk yang tidak berpengalaman. Untuk dapat mengetahui nilai *Entropy*, maka dapat menggunakan perhitungan dibawah ini:

Entropy (Berpengalaman) :

$$= ((-478/695)*\text{Log}_2(478/695)+(-217/695)*\text{Log}_2(217/695)) \\ = 0.895731108$$

Entropy (Tidak Berpengalaman) :

$$= ((-173/265)*\text{Log}_2(173/265)+(-92/265)*\text{Log}_2(92/265)) \\ = 0.931515027$$

Kemudian untuk mendapatkan nilai *Gain* nya, dapat menggunakan perhitungan sebagai berikut:

$$\text{Gain Pengalaman Kerja} = (0.906408808)-((95/960)*0.895731108)-((265/960)*0.931515027) \\ = 0.000799848$$

5. Atribut Tes 1

Untuk atribut Tes 1 ini, memiliki data sebanyak 506 untuk data baik, 251 untuk data cukup dan 203 untuk data yang kurang, untuk mengetahui nilai *Entropy* dari atribut ini, maka dapat dilakukan perhitungan sebagai berikut :

Entropy (Baik) :

$$= ((-496/506)*\text{Log}_2(496/506)+(-10/506)*\text{Log}_2(10/506)) = 0.140106913$$

Entropy (Cukup) :

$$= ((-94/251)*\text{Log}_2(94/251)+(-157/251)*\text{Log}_2(157/251)) \\ = 0.954066225$$

Entropy (Kurang) :

$$= ((-61/203)*\text{Log}_2(61/203)+(-142/203)*\text{Log}_2(142/203)) \\ = 0.88189223$$

Kemudian untuk dapat mengetahui hasil nilai *Gain* nya dapat menggunakan perhitungan sebagai berikut :

$$\text{Gain Tes 1} = (0.906408808)-((506/960)*0.140106913)-((251/960)*0.954066225)- \\ ((203/960)*0.88189223) = 0.396628763$$

6. Atribut Tes 2

Untuk atribut Tes 2, memiliki data sebanyak 418 untuk data baik, 410 untuk data cukup dan 132 untuk data yang kurang, untuk mengetahui nilai *Entropy* dari atribut ini, maka dapat dilakukan perhitungan sebagai berikut :

Entropy (Baik) :

$$= ((-418/418)*\text{Log}_2(418/418)+(-0/418)*\text{Log}_2(0/418)) = 0$$

Entropy (Cukup) :

$$= ((-220/410)*\text{Log}_2(220/410)+(-190/410)*\text{Log}_2(190/410)) \\ = 0.996134484$$

Entropy (Kurang) :

$$= ((-13/132)*\text{Log}_2(13/132)+(-119/132)*\text{Log}_2(119/132)) \\ = 0.464174194$$

Kemudian untuk mendapatkan hasil nilai *Gain*, dapat menggunakan perhitungan sebagai berikut:

$$\text{Gain Tes 2} = (0.906408808)-((418/960)*0)-((410/960)*0.996134484)- \\ ((132/960)*0.464174194) = 0.417152421$$

Evaluasi *Confusion Matrix* Algoritma Decision Tree (C4.5)

Aspek terakhir mengenai *confusion matrix* algoritma *decision tree* yang dievaluasi secara keseluruhan untuk mengukur seberapa baik memberikan nilai akurasi, presisi dan *recall*, berikut *confusion matrix* algoritma *decision tree* data testing di *rapid miner* antara lain:

	true DITERIMA	true TIDAK DITERIMA	class precision
pred. DITERIMA	128	1	99.22%
pred. TIDAK DITERIMA	2	61	96.83%
class recall	98.46%	98.39%	

Gambar 5 Evaluasi Confusion Matrix Algoritma Decision Tree

Dari proses evaluasi model *Decision Tree* dan sebuah proses validasi terbentuk dari hasil matrix *Accuracy* sebesar 98.44%, *Precision* sebesar 99.22% dan *Recall* sebesar 98.46%. namun hasil ini dapat diperhitungkan secara manual, berikut perhitungan manual akurasi, presisi dan *recall* dibawah ini.

$$Precision = \frac{Tp}{Tp+Fp} \times 100 \dots\dots\dots(3)$$

$$Recall = \frac{Tp}{Tp+Fn} \times 100 \dots\dots\dots(4)$$

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \times 100 \dots\dots\dots(5)$$

$$Accuracy = \frac{128+61}{128+61+1+2} = \frac{189}{192} \times 100 = 98.44\%$$

$$Precision = \frac{128}{128+1} = \frac{128}{129} \times 100 = 99.22\%$$

$$Recall = \frac{128}{128+2} = \frac{128}{130} \times 100 = 98.46\%$$

5. KESIMPULAN DAN SARAN

Proses yang sudah dilakukan terhadap algoritma *decision tree* (C4.5) dapat disimpulkan bahwa, hasil perhitungan *confusion matrix* diketahui akurasi tertinggi yang diperoleh 98,44% sehingga proses klasifikasi yang dilakukan untuk faktor kelayakan karyawan baru dilihat berdasarkan hasil tes 2.

6. DAFTAR REFERENSI

- Abdul Fath Ramadhan, R. A. S., & L. M. F. A. (2024). Penerapan algoritma C4.5 untuk penerimaan calon pegawai negeri sipil KEMENKUMHAM Sulawesi Tenggara. *Jurnal XYZ*, 14(2), 133–138.
- Abidin, Z., Nurhana, E., & Faruq, U. (2023). Analisis perbandingan algoritma decision tree C4.5 dan C50 pada data karyawan berpotensi promosi jabatan. *ZA Pagar Alam*, 17(2), 567–582.
- Adria, H. M. (2022). Analisis perbandingan aplikasi data mining dalam memprediksi kualitas kinerja karyawan menggunakan metode algoritma C4.5. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 9(2), 1655–1665. <https://doi.org/10.35957/jatisi.v9i2.1992>
- Ananda, R., Sumarno, S., & Gunawan, I. (2022). Prediksi promosi jabatan karyawan JNE Pematangsiantar dengan algoritma C4.5. *Jurnal Sistem Komputer dan Informatika (JSON)*, 3(3), 358. <https://doi.org/10.30865/json.v3i3.3763>
- Asri, A., Arifin, A., & Handoko, W. (2022). Implementasi metode Naive Bayes untuk klasifikasi penerima program keluarga harapan. *Journal of Computer*, 2(1), 21–26.
- Gina Sonia, & Indriyani, A. (2023). Analisis penerapan data mining dengan metode algoritma C4.5 untuk pendataan karyawan tetap di KONI Sumatera Utara. *Journal of Computer Science and Informatics Engineering (CoSIE)*, 2(1), 48–55. <https://doi.org/10.55537/cosie.v2i1.523>
- Hidayah, A. N. Z., & Rozi, A. F. (2021). Penerapan data mining dalam menentukan kinerja karyawan terbaik dengan menggunakan metode algoritma C4.5 (Studi kasus: Universitas Mercu Buana Yogyakarta). *Jurnal Information System & Artificial Intelligence*, 1(2), 117–127.
- Husein, A. M., & Brutu, M. (2022). Prediksi penerimaan calon karyawan dengan menggunakan algoritma C4.5 pada Biro Kesejahteraan Rakyat Provinsi Sumatera Utara. *Digital Transformation Technology*, 2(1), 16–20. <https://doi.org/10.47709/digitech.v2i1.1769>
- Khoeri, I., & Iskandar Mulyana, D. (2021). Implementasi machine learning dengan decision tree algoritma C4.5 dalam penerimaan karyawan baru pada PT Gitareksa Dinamika Jakarta. *Jurnal Sosial Teknologi*, 1(7), 615–623. <https://doi.org/10.59188/jurnalsostech.v1i7.126>
- Leidiyana, H., & Permana, A. A. (2020). Pemodelan klasifikasi dalam meningkatkan proses pemilihan calon karyawan dengan metode C4.5 dan jaringan syaraf tiruan. *JIKA (Jurnal Informatika)*, 4(1), 7. <https://doi.org/10.31000/jika.v4i1.2392>
- Purnama, A. I., Aziz, A., & Wiguna, A. S. (2020). Penerapan data mining untuk mengklasifikasi penerima bantuan PKH Desa Wae Jare menggunakan metode Naïve Bayes. *Kurawal: Jurnal Teknologi, Informasi dan Industri*, 3(2), 173–180. <https://doi.org/10.33479/kurawal.v3i2.348>

- Putri, H., Purnamasari, A. I., Dikananda, A. R., Nurdiawan, O., & Anwar, S. (2021). Penerima manfaat bantuan non tunai kartu keluarga sejahtera menggunakan metode Naïve Bayes dan KNN. *Building of Informatics, Technology and Science (BITS)*, 3(3), 331–337. <https://doi.org/10.47065/bits.v3i3.1093>
- Rahmawan, H. (2020). Penentuan rekomendasi pelatihan pengembangan diri bagi pegawai negeri sipil menggunakan algoritma C4.5 dengan principal component analysis dan diskritisasi. *Jurnal Tekno Kompak*, 14(1), 5. <https://doi.org/10.33365/jtk.v14i1.531>
- Romli, I., & Zy, A. T. (2020). Penentuan jadwal overtime dengan klasifikasi data karyawan menggunakan algoritma C4.5. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 4(2), 694–702.
- Rudianto, R., Wahyudi, T., & Handayani, P. (2022). Perbandingan akurasi C4.5 dan Naive Bayes untuk evaluasi kinerja karyawan PT Catur Sentosa Adiprana. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 5(2), 61–66. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2350>
- Saputra, R., Wijaya, D., Wijaya, T., & ... (2023). Penerapan data mining dalam proses penerimaan karyawan baru pada perusahaan outsourcing (SDM) dengan metode C4.5 menggunakan aplikasi Weka. *Jurnal Bidang XYZ*, 89–98. Retrieved from <https://ejournal.kreatifcemerlang.id/index.php/jbpi/article/view/101>
- Siallagan, E., Parlina, I., & Suhendri, D. (2022). Model aturan tingkat kepuasan pasien terhadap pelayanan Puskesmas menggunakan algoritma C4.5. *Zahra*, 1(2), 65–72.
- Tulus, T. H. L., Id Hadiana, A., & Santikarama, I. (2022). Sistem prediksi awal terhadap atrisi karyawan menggunakan algoritma C4.5. *Informatics and Digital Expert (INDEX)*, 4(1), 18–24. <https://doi.org/10.36423/index.v4i1.882>